ECON 4160, autumn term 2015. Lecture 6

Automatic model selection. Markov switching models

Ragnar Nymoen

Department of Economics

21 Sep 2015

References I

- Automatic model selection:HN 19
- Lecture 1b, and HN Ch. 7.6 and 9.5, about the cumulation of Type-1 error probabilities as a result of repeated testing

Automatizised Gets I

- A long controversy in econometrics (not only time series) is whether it is "best" to go **specific-to-general**, or whether **general-to-specific** is better.
  - Specific-to-general**:** start with a small (specific) model and enlarge it if it fails residual misspecification tests (on constancy tests)
  - General-to-specific (Gets): start with a large model (DH, call it *general unrestricted model*, GUM) and reduce it to a small one using statistical tests.
- Working with practical econometric modelling since 1984 I have spent a lot of time doing both specific-to-general and Gets. Mostly being very *inefficient* in producing models that have practical "use-value",

Automatizised Gets II

- Would be great to be more efficient. Not surprising that we now are offered computer programs for model selection that promise make us more efficient.
- *OxMetrics* includes one such program called *Autometrics*. It is the big brother of the *PcGets* program that HN talk about in Ch 19.
- The aim of Autometrics is to search for relevant variables with a high probability of success, at a low *cost of search* (low probability of including many irrelevant variables).
- Present briefly some of the main concepts that relate to Gets modelling and its implementation in Autometrics.
- After the lecture, you can start testing out *Autometrics* and see if it is useful for you.

Automatizised Gets III

- See the *PcGive Vol I* (posted on the webpage) on the for tutorial and documentation.

eq# **HN**		**Concept**	**Based on**
(19.1.1)	$Y_t = \sum_{i=1}^{N} \gamma_i Z_{i,t} + u_t$ $N \geq n$	GUM	DGP
(19.1.2)	$Y_t = \sum_{j=1}^{n} \beta_j Z_{j,t} + \epsilon_t$	DGP	Theory
(19.1.3)	$Y_t = \sum_{r=1}^{m} \delta_r Z_{r,t} + \eta_t$	Selected Model	Manual or automatic GETS

- The DGP is unknown, nevertheless a premise for having a possibility of finding it, is that we are able to formulate a GUM that contains the DPG.

- Hence, for model selection to work, the specification of the GUM cannot be “data-based”, Instead based on research **purpose**, **theory** and **existing models**
- Given the that the DGP is “in the GUM”, the aim is to select variables in such a way that (with a high probability):
  - $m = n$ Right number of variables. (Even this can be of value, think of the $n = 1$, $N = 20$ case for example)
  - $\hat{\delta}_r \approx \beta_j,\ r = j$ for $r = 1, \dots, n$ selected variables

Success criteria for search algorithms I

1. Be good at removing irrelevant variables
2. Be good at keeping relevant variables

- A first objection to Gets is that 1. is impossible to meet, because *multiple testing* will lead to i*nflated Type-1 error probability level*.
  - Unavoidable that the true significance level of selecting form GUM, call it $\alpha^{N \to m}$, usually will be larger than nominal (e.g. 5 %) significance level, α. See end of Lecture 1a. Call this *cost of search*
  - Cost of search need not (logically) be very high, even when *n* is large. Example: 100 orthogonal regressors: Then need only one test to select model!

Success criteria for search algorithms II

- In practice, multiple testing occurs: Argument for choosing low α (to keep $\alpha^{N \to m}$ down)
- Can get an impression of algorithm "quality" by Monte Carlo simulation.
- High ability to remove irrelevant variables is then measured by *gauge* $\approx$ the chosen α (which the program call the *target* level of significance.

- Finding algorithms that are good at keeping variables that matter has proven to be more difficult
  - How good an algorithm is can also be studied by Monte Carlo, and measured by *potency*. Ideal is *potency* = *100*

Example I

Assume that we formulate a GUM of the ADL type with two regressors and 4 lags:

$$Y_t = \beta_0 + \phi_i \sum_{i=0}^{4} Y_{t-1-i} + \beta_{1i} \sum_{i=0}^{4} Z_{1t-i} + \beta_{1i} \sum_{i=0}^{4} Z_{2t-i} + u_t \quad \text{(GUM)}$$

We can use PcNaive to see how Autometrics selects when the premise is that the DGP is

$$\begin{aligned} Y_t = {} & 1.2 Y_{t-1} - 0.5 Y_{t-2} + 0 Z_{1t} + 0.8 Z_{1t-1} + 0 Z_{1t-2} \\ & + 0.5 Z_{2t} + 0 Z_{2t-1} - 0.50 Z_{1t-2} + \epsilon_t \end{aligned}$$

Z_{1t} and Z_{2t} are generated by a VAR(1). The Π matrix is

$$\Pi = \begin{pmatrix} 0.2 & 0.2 \\ -0.3 & 0.8 \end{pmatrix}$$

ϵ_t is Normal(0,1). Is this DGP inside the GUM?

Markow Switching models I

- After the "oblig" we are going to consider non-stationarity in a systematic way.
- So called Markov Switching models represent a way of introducing two or two equilibria in the ADL models we use.
- For simplicity consider the ADL(1,1)

$$Y_t = \phi_0(S_t) + \phi_1 Y_{t-1} + \beta_0 X_t + \epsilon_t, \, \epsilon \sim N(0\,,\sigma^2) \qquad (1)$$

where S_t is a random variable for the regime that the system is in at time t.

Markow Switching models II

- If there are two regimes 0 and 1, we could also write:

$$Y_t = \phi_0(0) + \phi_1 Y_{t-1} + \beta_0 X_t + \epsilon_t$$
$$Y_t = \phi_0(1) + \phi_1 Y_{t-1} + \beta_0 X_t + \epsilon_t$$

  which shows the regime dependency of the intercept more clearly.
- Is $E(Y_t)$ time dependent, of regime-dependent?
- Since S_t is a random variable we can write down the probabilities of being in a regime,given the entire history indicted by $\mathcal{I}_t$

$$P(S_t = 0 \mid \mathcal{I}_t) \tag{2}$$
$$P(S_t = 1 \mid \mathcal{I}_t) = 1 - P(S_t = 0 \mid \mathcal{I}_t) \tag{3}$$

Markow Switching models III

- The Markov Switching module in the program allows us to estimate equations like (1) as well as the transitions probabilities (2) and (3)
- Give an illustration at the end.